Authority and Harm

Parry, Jonathan

Authority and Harm[1]

This paper explores the connections between two central topics in moral and political philosophy: the moral legitimacy of authority and the ethics of causing harm. Each of these has been extensively discussed in isolation, but relatively little work has considered the implications of certain views about authority for theories of permissible harming, and *vice versa*.[2] As I aim to show, reflection on the relationship between these two topics reveals that certain common views about, respectively, the justification of harm and the moral limits of authority require revision. The paper proceeds as follows. Sections 1 and 2 clarify the question to be addressed and set out two main claims that I will defend. Sections 3-5 argue for the first claim. Sections 6-9 defend the second. Section 10 concludes.

1. The Central Question

The core concern within the ethics of harm is obvious. Though harming others is normally morally prohibited, under certain conditions it is intuitively permissible, or even required. Theories of harm aim to provide a systematic account of the factors that determine when these exceptions arise.

The theorist of authority, by contrast, is concerned with the fact that certain persons and institutions claim to possess the moral power to issue commands and, by doing so, place others under obligations to act in certain ways. Paradigmatic examples include a parent directing their child to 'Clean up your room!', a policewoman ordering a car driver to 'Stop right there!', and a colonel commanding his troops to 'Hold your positions!'. At a more general level, states and legal systems claim to create obligations by enacting laws and through the pronouncements of officials. In all these cases, the commander purports to create new 'content-independent' reasons for action, over and above the subject's pre-existing reasons for and against the action commanded. Furthermore, these new reasons claim a privileged status in the subject's practical deliberation. They are not simply to be weighed alongside all her pre-existing reasons, but are instead intended to silence or 'pre-empt' (at least some of) those reasons, preventing them from bearing on how she now ought to act.

Despite the ubiquity of authority claims, there is a clear puzzle as to how they could be true. Put simply: How can I acquire something as morally serious as an obligation just by someone communicating her intention that it be the case?[3] A theory of authority then faces two closely related tasks. The first is to identify the conditions, if any, under which this power is morally justified and obedience therefore required. The second is to provide an account of its moral limits, since obligations to obey are presumably neither unconditional nor absolute.

Despite their different objects of justification – harm vs. obedience – these two topics are ultimately concerned with what moral reasons agents have; with what individuals all-things-considered ought and ought not to do. Given this, there is a range of cases in which answering these questions requires determining how our accounts of harm and of authority interact with one another. These are cases in which an authority's command requires its subject to cause, or refraining from causing, harm to others.

Under what conditions, if any, do these commands give subjects all-things-considered reason to obey? This is no hypothetical question. For example, members of law enforcement and military organizations are routinely subject to such commands. The question is most striking in the case of commands to perform acts of harming that would be morally prohibited in the absence of the command. Here, two putative sources of obligations require opposing actions. A theory of authority or harm will be incomplete unless it tells us how conflicts like this should be resolved, by providing an account of the extent (if any) to which authoritative commands can affect the moral status of harmful actions.

2. Two Claims

To demonstrate the relevance of authority I focus on two specific debates within the ethics of harm. The first is the very general question of identifying the *range* of considerations that are capable of justifying harm. On a fairly standard view, the stringent constraint on harming is explained in terms of individuals having basic rights against harm. Given this, justifications for harming are thought to take one of two basic forms. The first is that the individual harmed *lacks* their normal right against harm, and so harming them does not wrong them. For example, they may have waived their right (as in the case of boxing matches), forfeited their right in virtue of some prior wrongdoing (as in the case of punishment), or rendered themselves liable to harm in virtue of posing an unjust threat to others (as in cases of self- and other-defense). A second form of justification holds that individuals' rights not to be harmed can be *overridden* by weightier moral reasons. Most obviously, that harming a person directly prevents a much greater harm to others. In these cases, harm is justified as the (impartial) lesser-evil.[4]

The above are often classified as agent-neutral justifications, in that they do not make essential reference to any particular agents to whom they apply.[5] For example, if John is liable to defensive killing, or if killing John will save many innocent lives, then any agent may potentially act on these justifications. In addition, some theorists defend the existence of agent-relative justifications, which apply only to particular agents.[6] These are typically grounded in considerations of permissible partiality. On this view, when certain agent-relative reasons – such as protecting oneself and one's loved ones – come into conflict with respecting others' rights not to be harmed, the agent-relative reason may sometimes be weightier. Given its structure, this can be understood as a distinct species of lesser-evil justification (Lazar, 2013).

While the precise range of justifications for harming is much debated, it is generally assumed that the above candidates exhaust the possibilities.[7] Term this view *Completeness*.

The second debate arises within discussions of defensive harm. It concerns the permissibility of using defensive force against individuals who threaten unjust harm to others, but possess an all-things-considered justification for doing so. This issue rests on a more general question regarding how different agents' reasons for causing harming *interact* with one another. While it seems plausible that an agent's justification gives others *some* reason not to defensively harm them, debate centers on the extent to which

the justificatory burden is raised. On one prominent (but by no means universal) view there can be no justified defense against justified infringements of rights against harm.[8] Term this view *Immunity*.

I will argue that extending a certain view of the justification of authority into the domain of harm generates counter-examples to both *Completeness* and *Immunity*. To do so, I defend two specific claims. Firstly, I defend the strong claim that, under certain conditions, the command of an authority can provide an agent with a moral justification for causing harm, even in cases where the harm both transgresses rights and fails to bring about goods sufficient to override those rights. This claim thus denies *Completeness*, positing an additional 'authority-based' form of justification.[9]

With the first claim in place, I shift from the question of the normative situation of those *subject* to commands to cause harm, to those who are *threatened* with harm by authorized agents. In particular, I consider the permissibility of defensively harming such agents. I defend a second claim, which holds that an agent's having an authority-based justification for harming does not, in itself, raise the justificatory burden on defensively harming that agent, compared to if they lacked that justification. This claim thus denies *Immunity*.

3. Opposing the First Claim

I anticipate many will find my first claim highly unintuitive, even repugnant. To begin a defense of this claim, I will outline three broad views about authority that support this common-sense reaction.

At the most general level, one might endorse philosophical anarchism and deny that the commands of authorities *ever* create reasons for action. This challenge is often stated in the form of a paradox, starting from the plausible assumption that agents should always act in accordance with the balance of reasons that apply to them.[10] Given this, in cases where an authority commands acting against the balance of reasons, obeying the command seems to involve acting against reason. If, on the other hand, we are commanded to act as reason recommends, then we ought to do so, but not *because* we have been commanded. From the perspective of practical reason, commands are either redundant or pernicious.

However, endorsing anarchism simply in order to resist my first claim does seem a case of killing the baby to save the bathwater. Fortunately, a more moderate, and plausible, strategy is available. This accepts that there is some successful response to the anarchist's challenge – so that *some* authorities are capable of creating *some* obligations – and instead appeals to the moral limits of that power. This is a very natural position to take. It seems obvious that wherever the precise limits lie, commands to cause harm that would otherwise be morally unjustified surely exceed them, given the gravity of the wrongdoing involved. As Matthew Noah Smith puts it in a recent article,

> The first characteristic of the obligation to obey the law is that there are very few limits on what an obligation to obey the law can require a subject to do. There are, of course, some limits. *Presumably, if obedience to the law requires commission of serious moral wrongs, then one is not obligated*

> *to obey the law*. But this limit is at the moral extremes." (Smith, 2013, p.349) (my emphasis)

This common thought supports two distinct objections to my first claim, corresponding to two different ways in which the obligation to obey a legitimate authority is limited.[11] The *Invalidation Objection* holds that the obligation to obey is necessarily *voided* when the authority's commands require actions that would otherwise be seriously morally wrong.[12] These commands create no reasons to obey. By contrast, the *Pro Tanto Objection* grants that such commands may succeed in creating obligations, but holds that these obligations are necessarily *overridden* by the subject's weightier duty not to cause serious harm to others.[13]

3. Service Justifications of Authority

In order to defend my first claim, a plausible account of authority is needed that reveals all three objections to be mistaken. This requires two components. Firstly, in response to the anarchist, we need an account of how one person's authority over another – understood as the moral power to create content-independent and peremptory obligations – can be morally justified. Secondly, in response to the Invalidation and *Pro Tanto* objections, it needs to be shown that commands to inflict unjustified harm need not necessarily exceed the moral limits of authority. I will argue that 'service' accounts of authority are able to satisfy both these requirements. On this view, very roughly, one agent's having authority over another can be justified when, and to the extent that, the authority having this moral power serves the subject's ends.

Let me begin by outlining Joseph Raz's well-known argument for justifying authority in this way (Raz, 1986). This advances two main theses. The first ('pre-emption') thesis explains the peremptory character of commands in terms of a hierarchical account of practical reasons. On this view, an authoritative command to φ is intended to give its subject both an additional first-order reason for φ-ing and a second-order *exclusionary* reason not to act on the basis of (some of) the pre-existing first-order φ–related reasons. These reasons are supplanted by the command.

Of course, the fact that commands are intended to play this role does not show that they do so. This second-step is provided by the second ('normal justification') thesis, according to which,

> the normal way to establish that a person has authority over another person involves showing that the alleged subject is likely better to comply with reasons which apply to him . . . if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons which apply to him directly. (Raz, 1986, p.53)

On this view, authority is justified in virtue of the rational gains it provides its subject. An authority is entitled to create new obligations by issuing commands because, and to the extent that, its having this ability the subject to better achieve aims they have

independent reason to achieve.[14] Authorities can satisfy this test in two main ways. Firstly, obeying a common authority may enable individuals to better coordinate their behavior with one another, thereby resolving various collective action problems they may encounter in pursuing valuable aims. Secondly, an authority may possess greater expertise than the subject on certain morally important matters (where expertise is understood broadly, as the ability to issue directives that track right reason more reliably or efficiently than the subject is able to.)[15]

The normal justification thesis thus offers a broadly instrumental account of authority, thereby responding to the anarchist's worry that obedience necessarily conflicts with reason.[16] Obeying an authority may simply be the optimal means of achieving one's ends and, when so, obedience is justified (Raz, 2010, p.299).[17] This idea also explains the pre-emptive character of authoritative commands: The subject best conforms to reason by allowing commands to replace their own practical assessment of certain considerations. It is easiest to illustrate this point in cases of expertise-based authorities (another, though broadly parallel, story has to be told with respect to coordinative authorities.) To put things somewhat crudely, such authorities are less likely than the subject to make mistakes as to what reason requires within an identifiable class of cases. Under these conditions, if the subject assigns its commands a pre-emptive role, she will achieve an identical level of success that the authority achieves. Alternatively, she could adopt a non-pre-emptive strategy, in which she simply gives the reasons that favor the action commanded some additional additional weight in her deliberations. In a sub-class of cases, this weight will tip the balance in favor of acting as commanded, and her rate of mistake will match the authority's. In the remaining cases the command will not tip the balance and she will act according to her own assessment. Here, her rate of mistake will exceed the authority's. Across the total class, then, the subject does worse than the authority. A weighing strategy can only serve to reduce her overall conformity with reason, compared to preemption. Instrumental reason thus dictates that commands have preemptive force (Raz, 1986, pp.67-69).[18]

A service-based view also explains why mistaken commands – commands that fail to reflect the balance of reasons in a particular case – can still succeed in creating obligations. This is because authorities do not need to be infallible in order to serve their subjects. Provided the authority is better placed that the subject with respect to achieving conformity with reason, the subject still improves their overall performance by obeying. Crucially, subjects can only gain the benefits of authority if its commands remain binding even in certain cases where they fail to track right reason. For, in order to avoid acting against reason in such cases, the subject would have to rely on their own assessment of the relevant considerations. But such a policy requires forsaking the overall gains of obedience, since the authority meets the condition of normal justification despite its fallibility.

To demonstrate, consider a simple case of advisory (rather than practical) authority, in which A has authority over B within the domain of financial investment. B will overall better maximize his returns by following A's directives, rather than by acting on his own judgment. This is compatible with A, from time to time, mistakenly

directing B to make poor investments**,** costing B one hundred dollars each time. However, detecting these mistakes would require engaging in the same process of financial reasoning that B went through in each case. If B does so, and acts on his own judgment, he will overall do worse in terms of maximizing his returns, compared to a more general policy of obedience. B therefore has sufficient reason to act as A directs, including in cases where A errs.

However, this doesn't mean that all commands from a legitimate authority create reasons to obey. The validity of commands is limited in two respects on a service-based view (I discuss their limited weight in Section X). The first restricts the jurisdiction of authority. Given the value of autonomy, there will be a range of domains in which agents' overriding rational aim is to choose for themselves, rather than achieve the 'optimal' outcome. For example, one's choice of leisure activity, romantic partner, religious affiliation, etc. Given the priority of autonomous choice in these areas, obeying an authority would be self-defeating. Such domains are just not 'authority-apt' and commands issued within them are void.[19]

Service accounts also limit authority at the level of specific commands, as well as domains. Though directives may remain binding even if they fail to reflect right reason, this doesn't mean that *all* mistaken directives bind. Service justifies obedience only to the extent necessary to optimize the subject's overall conformity with reason. Commands that require obedience beyond this point are invalid. When disregarding a mistaken command does not incur a rational cost, the subject is free, in fact required, to do so. To illustrate, consider a variation on the financial advisor example, in which A mistakenly directs B to burn ten of his dollars. In this case, B can judge that the directive is mistaken without having to engage in any complex financial reasoning of the kind that A is superior to B at doing. He can therefore disregard it without forfeiting the benefits of generally following A's directives. To clarify, whether a command's departure from right reason serves to invalidate it does not depend on how large a mistake it is. In our pair of financial examples, B's conforming to the first directive loses him ten times as much money as conforming to the second. Yet only the first is reason-giving. Instead, validity depends on the *type* of mistake. As Raz (1986, p.62) puts it, what matters is the *clarity* of a mistake, not its *gravity*. Only clear mistakes invalidate, because only disobeying clearly mistaken commands is compatible with optimizing one's conformity with reason.[20]

4. The Authority View of Harm

My contention is that if we accept a service account of how the commands of authorities can *ever* create obligations, it is a relatively short step to accepting that the commands of authorities can give subjects sufficient reason to cause otherwise-unjustified harm, thus vindicating my first claim. I set out this argument below and discuss some of its intricacies in the next section.

A service-based view provides a very general account of the justification of authority: A has authority over B within domain X, if obeying A's commands enables B to better conform to the X-related reasons that apply to B. The argument from this general account to a defense of my first claim proceeds in four steps.

The first is to make one element of this three-place relation more specific: the domain of authority. Presumably, unless extreme pacifism is true, there are possible domains in which acting in accordance with reason may involve causing harm to others. Term such domains *harm-apt*. I mentioned two possible examples earlier – the domains of military service and law enforcement.

The second step simply notes that agents operating in harm-apt domains may be differently situated regarding their abilities to assess and successfully bring about conformity with the harm-related reasons. Term this 'agent-variability'.

The third step combines the first two. Harm-aptness and agent-variability open up the possibility that agents may better conform with the harm-related reasons by obeying the commands of another, rather than by trying to conform to those reasons directly. This shows how one agent may acquire authority over another regarding the distribution of harm. Put differently, domains can be both harm-apt and *authority-apt*. If an agent will better distribute harm by obeying the commands of an authority, it seems uncontroversial that this is what they all-things-considered ought to do.

A fourth and final step is required to support my first claim. This is provided by the fact that, as explained above, authorities can be legitimate despite their fallibility. Subjects can be all-things-considered required to obey commands that fail to reflect right reason. When an authority serves its subjects within a harm-apt domain binding, yet mistaken, commands may include those that require distributing harm in ways that are not justified on the basis of the authority-independent reasons.

Term this four-stage argument the *Authority View of Harm.* To illustrate it, consider the following example:

Volcano: A volcano erupts in Nation A. In order to save as many lives as possible the lava flow needs to be diverted from areas of higher population density to lower. This requires Nation A's citizens to dig an integrated system of trenches, along which the lava can be redirected.

Assume that Nation A's citizens will do better in terms of saving lives by obeying their government on matters of lava-redirection, compared to not obeying. This may be due to the government's ability to achieve coordination among its subjects (because whether any individual trench-digger contributes to successful lava redirection depends on what other trench-diggers do), or its expertise (it makes sufficiently good decisions regarding lava redirection), or a combination of both. According to the Authority View of Harm, Nation A's government thereby acquires authority over its subjects regarding the domain of lava redirection. Nation A's citizens have a duty to obey their government on matters of lava redirection, including certain commands that are mistaken and require harming innocents in the absence of a lesser-evil justification. Since this policy of obedience is their optimal means of distributing harm, they are morally required, all things considered, to do so.

In summary: I have argued that if a broadly service-based view is defensible, a common and intuitive view about the moral limitation of authority is mistaken. It is not true, as a general matter, that commands to perform seriously wrongful actions

necessarily fail to generate all-things-considered obligations to obey. This result also denies the common assumption that justifications for harm fall into one of two categories, in which the reasons against harming are either vitiated or overridden. Instead, there exists an additional form of justification, in which these reasons are defeated by *exclusion*.

Before moving on, it is worth considering an important objection to the Authority View.[21] The objection holds that the subject's reasons to obey are not of the right sort to justify causing harm. More specifically, it claims that if the reasons to obey an authority arise from its superior expertise, its commands only provide the subject with reasons to *believe* that their actions are justified, and not practical reasons to *act* as commanded. Hence, subjects are 'justified' in causing harm only in the 'evidence-relative' sense, which may furnish them with an excuse for harming, but not a moral permission. This specific worry echoes a more general objection that service accounts can only establish epistemic, and not practical, authority (see, for example, Darwall, 2010).

In response, it is not obvious that the reasons to obey an expertise-based authority can be straightforwardly reduced to reasons for belief, as the critic claims. This is because the subject's epistemic aim of forming true beliefs about the world can come apart from her practical aim of improving their conformity with reason by obeying an expertise-based authority. For example, it may be that within an identifiable range of cases, the subject will more successfully form true beliefs about the balance of reasons in each case by rely solely on her own assessment, compared to deferring to the authority. Yet she may still do better in terms of conforming her *behavior* with right reason by obeying the very same authority. This could be because, although she may make fewer mistakes than the authority, the mistakes that she does make are more serious, such that *acting* on her assessments will yield worse practical results than a policy of obeying the authority, including (at least some of) its mistaken commands. Cases such as this suggest that one can have reason to obey an expertise-based authority, even if its directives do not provide reasons for forming beliefs. If so, the commands of such authorities create reasons for action, and so are capable of justifying behavior (including harmful behavior) more robustly than the mere evidence-relative sense.

Though I am sympathetic to this line of response, more clearly needs to be said. Fortunately, a more straightforward response is also available. This simply points out that the objection has very limited scope, since it applies only in the case of authorities that are justified solely on the basis of expertise. But expertise is not the only, or even the main, way of justifying authority in terms of service. In many (perhaps most) cases, authorities are legitimated on the basis of their ability to enable their subjects to coordinate their actions with one another, so that they can better achieve morally important goals.[22] In the *Volcano* case, for example, whether each citizen successfully contributes to saving lives depends on coordinating their actions with others. The Authority View claims that if obeying an authority enables the required coordination, then the citizens are required to do so, including in (at least some) cases where the authority issues mistaken commands. For present purposes, the important point is that

reasons to coordinate are clearly practical reasons, and not merely reasons for belief. Given this, even if we concede that expertise-based authorities are merely epistemic authorities[23], this does not significantly undermine the Authority View. At most, it reduces the range of cases in which authority-based justifications for harm apply. However, given that right action in many paradigmatic harm-apt domains (such as law enforcement and military action) will depend on coordination, this doesn't seem especially troubling.

5. The Moderate Objections Revisited

This section refines the Authority View by explaining why the two moderate objections outlined above fail to refute my first claim. According to the *Pro Tanto* Objection, commands to cause unjustified harm are necessarily overridden by the duty not to transgress the basic rights of others. However, as the Authority View reveals, it is a mistake to treat all such cases in terms of a straightforward competition of reasons. In order for authorities to successfully serve their subjects, their commands must have the status of pre-emptive reasons, excluding the reasons on which they are based. This is equally true in harm-apt domains as in any other

However, this does not mean that valid commands cannot be overridden by weightier first-order reasons. This is perfectly admissible on a service account, provided that the reasons in question do not fall within the authority's jurisdiction (Raz, 2009, p.144-146). To demonstrate, imagine that Smith has the aim of acting rightly on some morally important matter. The correct course of action depends on a trade-off between three distinct variables, X, Y and Z. Furthermore, imagine that an authority passes the test of normal justification regarding Smith within the domains of the X-related and Y-related reasons, but not the Z-related reasons. Under these conditions the authority's command excludes variables X and Y from Smith's practical reasoning. But the command may perfectly permissibly be weighed against the Z-related reasons. Furthermore, it is entirely possible that the non-excluded Z-related reasons are sufficiently weighty to override the obligation created by the command, giving Smith all-things-considered reason to disobey.

Given this, the Authority View is compatible with there being cases in which the *Pro Tanto* Objection gives an accurate picture of the normative situation. For example, in some contexts distributing harms correctly may require a trade-off between minimizing harm and distributing it equitably. In these cases, an authority might successfully serve its subjects regarding the (sub)domain of harm minimization, but not the (sub)domain of equity. Like all of us, authorities are typically better at some things than others. Under these conditions, commands to cause (or refraining from causing) harm only exclude reasons pertaining to harm-minimization. Equity-based considerations are not excluded and, in some cases, may be sufficiently important to outweigh the obligation created by the command. In cases like this, the subject will have both an obligation to inflict unjustified harm and a weightier countervailing reason not to do so. However, the important point is that while there may be many cases that have this structure, it is not true of *all* cases. This is what the *Pro Tanto* Objection requires if it is to refute my first claim.

The Invalidation Objection claims that an authority's commands only create obligations if their content does not significantly depart from the balance of moral reasons. Authorities that issue such commands necessarily exceed their legitimacy. Hence, commands to impose harms (or at least sufficiently serious harms) that are not independently justified are void. The problem with this objection is that service accounts provide a very general model of how the moral power to create content-independent obligations can be justified, which applies across different domains of reasons. Given this, it is hard to find a principled rationale for the localized denial of this power that the objection requires. If the aim of improving one's conformity with reasons can *ever* explain why commands that require acting against the balance of pre-existing reasons create obligations, why should it not also do so regarding the reasons that govern the distribution of harm? It is arbitrary to simply carve off this domain as immune from a service-based justification.

However, this may be too quick. On a service-based view there are certain domains in which commands *are* necessarily and non-arbitrarily invalid: those in which choosing autonomously is more important than achieving optimal outcomes. One might then resurrect the Invalidation Objection by claiming that agents have more reason to distribute harms autonomously than optimally. In other words, harm-apt domains are never authority-apt.[24] If so, commands to inflict unjustified harm *would* necessarily fail to create obligations.[25] But this is very hard to believe. If there is any domain in which improving one's conformity with reason trumps the value of exercising autonomy, it is surely that of harm-distribution. Appealing to autonomy cannot rescue the Invalidation Objection from the charge of arbitrariness.[26]

The Invalidation Objection may also be revised in a different direction. As explained above, commands that are *clearly* mistaken create no reasons for action. Given this, one might argue that commands whose content seriously departs from the balance of moral reasons *also* constitute clear mistakes. This would provide a non-arbitrary basis for the claim that commands to cause unjustified harm are invalid in virtue of their immoral content, since whether a command constitutes a clear mistake *is* determined by its content. However, on this revised view it is the clarity of a command's departure from right reason, and not its immorality *per se*, that accounts for its invalidity.

However, it is highly implausible that every command to cause otherwise-unjustified harm also constitutes a clear mistake. In order for a command to qualify as a clear mistake, the subject must be able to determine that the command fails to reflect right reason without engaging in the same reasoning that the authority went through in producing its commands. Importantly, whether or not the subject can form such a judgment depends not only on the command's content, but also on the particular domain in which the command is issued and the nature of the service that the authority provides. Given this contextual element, the very same command may constitute a clear mistake when issued in one domain, but not in another. Perhaps *some* commands constitute clear mistakes across all domains, such as those that require impossible actions ('Do X and not-X! Now!'). But the claim that *every* command to inflict unjustified harm constitutes a clear mistake is surely false. Harm-apt domains are precisely those in which

determining that such commands are mistaken frequently (though not always) requires repeating the authority's deliberations.

Both the *Pro Tanto* and Invalidation Objections are thus unsuccessful in refuting my first claim. They do not fail because they misidentify ways in which authority is limited. Service accounts agree that the commands of authorities are limited in terms of both their weight and validity. Rather, they fail because they assume that the question of whether particular commands exceed those limits can be settled independently of a specific account of authority's justification. I have argued that this is a mistake. On a service-based view, the scope and limits of the obligation to obey are calibrated to what is required for the authority to provide the relevant service. When the service consists in enabling subjects to better distribute harm, subjects can be required, all-things-considered, to obey (at least some) commands to cause unjustified harm.

6. Defending The Second Claim

In what follows, I shift focus from the question the *range* of reasons that are capable of justifying harm, to that of how different agents' reasons for harming *interact*. More specifically, if my first claim is defensible and authoritative commands can provide an independent source of justification, to what extent does an agent's possession of an *authority-based* justification for causing unjust harm affect whether other agents are permitted to defensively harm the authorized agent? Whereas the preceding discussion centered on those who are *subject* to commands to cause harm, the following concerns the normative situation of those who are *threatened* by authorized agents.

To recapitulate, within the literature on defensive harm several theorists defend the view I labelled *Immunity*, which holds that there is no justified defense against justified infringements of rights against harm. In opposition, I argue that authority-based justifications reveal *Immunity* to be mistaken. The fact that an agent is justified in causing unjust harm *in virtue of being commanded* does not, in itself, raise the justificatory burden on defensively harming that agent. Though denying *Immunity* is not an uncommon position in itself, the argument I offer is distinctive because it is compatible with certain commitments that are often taken to strongly support *Immunity*.

Discussions of the permissibility of harming justified threateners typically focus on cases in which the threatener possesses an (impartial) lesser-evil justification for harming others. These provide a useful starting point for assessing the case of authority-based justifications. A standard test case is the following:

Tactical Bombers: A bomber crew are on a mission to destroy a munitions factory as part of a just war. Destroying the factory will result in the deaths of five innocent bystanders as a side-effect. However, the good achieved by bombing the factory is sufficient to justify doing so as the lesser evil. The five bystanders have access to an anti-aircraft gun and are able to shoot down the bombers before they drop their bombs.

The question here is whether the bystanders are permitted to defensively kill the bombers, given that the bombers are justified in causing their deaths. While many hold that the bystanders would be so permitted (Hosein, 2014; Mapel, 2010; Rodin, 2011;

Steinhoff, 2008), others argue that the bombers' justification entails that defense is impermissible (Frowe, 2015; McMahan, 2014; Tadros, 2011, ch.9).

The debate between these two views often turns on one's position on the *range* of justifications for harming. As explained above, justifications are standardly divided into agent-neutral and agent-relative. Agent-neutral justifications – such as defensive liability and (impartial) lesser-evil – apply to all agents, whereas agent-relative justifications – such as those grounded in permissible partiality – apply only to specific agents. If one takes the range of justifications to be thoroughly agent-neutral, then a commitment to *Immunity* follows quite naturally. If the reasons that determine how harm ought to be distributed in any particular case apply equally to all agents, then this gives every agent the common aim of seeing to it that that this distribution comes about, or at least not preventing it from coming about.[27] For the agent-neutralist, it is contradictory to hold that certain agents may be justified in bringing about one distribution of harm, while others are justified in bringing about an opposing distribution.[28]

Conversely, if one accepts the possibility of agent-relative justifications, then *Immunity* need not hold. If some forms of justification apply only to specific agents, there is no oddity in claiming that different parties can be simultaneously justified in harming one another. For example, while the bombers may possess a lesser-evil justification, the innocent bystanders may be justified in resisting on the basis of permissible self-partiality.

7. Authorized Threateners and Immunity

Let us now consider the permissibility of defense against authorized threateners. On first impression, it is tempting to endorse *Immunity* here and hold that the permissibility of violent resistance is precluded by their justification. This view is appealing because it generates the intuitively right result in a range of cases, such as the following:

Police Officer: A police officer acts to arrest an individual as a result of a command to do so from a morally justified authority. However, the command is mistaken (but not clearly so) and the prospective arrestee is innocent.

In this case, it seems impermissible for the arrestee to use defensive force against the police officer. Combining the Authority View of Harm with *Immunity* provides a neat explanation of why this is so. The Authority View allows us to characterize the police officer as posing a justified threat to the innocent arrestee, despite the fact that the harm is not justified by the command-independent reasons. The addition of *Immunity* allows us argue that the police officer's justification for harming defeats the arrestee's normal permission to resist aggression.[29] Furthermore, this analysis also yields the intuitively right result in a variation on the case:

Vigilante: A private individual acts to carry out a citizen's arrest on the basis of a reasonable suspicion that the arrestee will otherwise commit a serious crime. However, they are mistaken and the prospective arrestee is innocent.

In this case it *does* seem intuitively justified for the arrestee to forcefully resist. Again, combining the Authority View with *Immunity* neatly explains this. Although, by hypothesis, both the police officer and the vigilante threaten an identical harm, only the police officer possesses a justification for doing so, because only the police officer threatens harm in conformity with an authoritative command. Though the vigilante may reasonably believe that they are justified in harming the arrestee, they in fact lack sufficient reason to do so. Since this analysis classifies the vigilante as a species of unjustified threatener, *Immunity* does not apply and resistance may then be justified (subject to the usual requirements of necessity and proportionality).

However, other cases strongly suggest that *Immunity* does not hold in the case of authority-based justifications. Consider the following:

Combatants: A group of combatants act to annex an area of territory belonging to a neighboring state as a result of legitimate command to do so. However, the command is mistaken (but not clearly so) and the invasion is unjustified.[30]

In this case it seems clearly permissible for those threatened by the authorized agents to resist (or for third-parties to do so on their behalf). Yet applying *Immunity* to this case generates the opposite result. Surrender would be morally required, which is highly counter-intuitive.[31]

The interaction question thus raises an important challenge for the idea of authority-based justifications, in the form of a dilemma. Since, by hypothesis, both the police officer and the combatants possess the same form of justification for harming, we cannot claim that *Immunity* applies to one but not the other. Either *Immunity* holds in both cases – giving the wrong result in the *Combatants* case – or fails to apply in both cases – giving the wrong result in the *Police Officer* case.

8. Authorized Threateners and Agent-Relativity

I propose an account of interaction for authority-based justifications that aims to avoid the dilemma. The proposal has two parts. First, I argue that *Immunity* does not apply in the case of authority-based justifications for harming, thereby avoiding the first horn. Second, I provide an alternative and non-*ad hoc* account of why defense may be unjustified in cases such as *Police Officer*, thus avoiding the second horn. This section defends the first part; the following section argues for the second.

Recall the above discussion of the relationship between views about the range of reasons that are capable of justifying harm and views about how those reasons interact interpersonally. Those who take the range of justifications to be thoroughly agent-neutral are typically committed to *Immunity* as an account of interaction, whereas those who accept the existence of agent-relative justifications deny it. Given this, one strategy for denying that *Immunity* applies to authority-based justifications is to argue that the reasons that constitute the latter are agent-relative. Since agent-relative reasons need not affect the normative situation of other agents, the possession of an agent-relative

justification for bringing about a certain distribution of harm does not mean that others also have reason to bring about that distribution.

The argument for this view is tentative, but fairly straightforward: authority-based justifications fit the standard characterization of agent-relative reasons. One can only give a full statement of the reason for action provided by a legitimate command by making explicit and ineliminable reference to a particular agent for whom it is a reason. When an authority issues a command to φ this is not intended to bring a new reason for φ-ing into existence for all agents generally, but only for the subject(s) of the command. Moreover, the agent-relativity of the reasons created by legitimate commands is particularly salient under service accounts of authority, which require the obligation to obey to be demonstrated anew with regard to each subject and their particular circumstances.[32] On the view that legitimate authorities are those that enable individuals to compensate for various deficiencies and shortfalls in their practical reasoning, the justification of an individual's obligation to obey must necessarily appeal to specific facts about *that* individual.

If this characterization of authority-based justifications as agent-relative is defensible, we have the beginnings of an explanation of why *Immunity* does not hold in cases such as *Combatants*, thus avoiding the first horn of our dilemma. Though, by hypothesis, the combatants possess sufficient reason for causing harm, these reasons do not affect the normative situation of others, and so do not count against resistance.

It may be objected that this claim is too strong.[33] The objection proceeds from the following assumption: that all agents have a *pro tanto* reason to promote others' conformity with reason, which includes enabling them to be served by authorities. Given this, one may claim that prospective victims (as well as third-parties) do in fact have *some* reason not to resist authorized threateners, because resistance would prevent the authorized agent from conforming to reasons that apply to them: those provided by their authority's command. Hence, I am mistaken to claim that the authorized-threatener's justification does not count against resistance on the part of their victims.

However, I don't think this conclusion follows from the assumption. It may well be true that I have reason to promote all other agents' conformity with reason, and that doing so may involve bringing it about that others *are subject* to authorities that serve them. But it does not follow from this that I necessarily have reason to promote others *obeying* authoritative commands, in cases where the command fails to reflect the balance of pre-existing reasons.[34] While *the subject* of the mistaken command may have sufficient reason to obey it, they do so only because a policy of obedience is an optimal, though imperfect, means *for them* to achieve greater overall conformity with their ultimate reasons. When the strategy goes awry in particular cases, I should be guided by the subject's ultimate reasons, not their instrumental reasons. Hence, in cases like *Combatants*, the victims' aim of promoting others' conformity with reason does not give them reason to refrain from resisting their attackers. If anything, it gives them reason to resist.

Though the preceding points go some way towards showing why defensively harming authorized-threateners can be justified, but they are not yet sufficient for this

conclusion.[35] While they may show that the *reasons that justify* the authorized threatener do not count against violently resisting them, it might still be the case that the fact *that the threatener is justified* does so. More specifically, it might be objected that the authorized threatener's justification exempts them from liability to defensive harm. If so, this would impose a significant, perhaps even decisive, constraint on harming them.

The first point to note in response is that the doctrine 'justification defeats liability' is controversial.[36] One potential problem is that justification does not typically defeat other kinds of moral liability, such as liability to compensate *ex post* for causing harm.[37] The second, more substantial, point is that (as far as I am aware) those who endorse the doctrine have only explicitly defended it with respect to standard cases of impartial lesser-evil justifications, such as the *Tactical Bombers*. We should therefore be cautious in claiming that these arguments generalize to other forms of justification, and to authority-based justifications in particular. In fact, there are reasons to doubt that they do. Take, for example, the most sustained defense of the doctrine, put forward by Jeff McMahan. On McMahan's view, the doctrine is grounded in a specific account of the basis of liability, according to which "the assignment of liability follows the distribution of harm in accordance with the demands of justice." (McMahan, 2008, p.234). In the case of standard justified threateners, who have an impartial lesser-evil justification, "there is no reason that justice would demand that unavoidable harm be distributed towards them" (p.234), and so they are exempt from liability. While I find this view quite plausible, the rationale clearly does not apply to agents who possess authority-based justifications, since precisely what these justify is acting *contrary to* the just distribution of harm.[38] By the lights of McMahan's account, authorized threateners should be liable to defensive harm. What this shows, I think, is that justification *per se* does not defeat liability (if indeed it defeats it at all). Rather, it depends on the kind of reasons that provide the particular justification.

This argument regarding liability completes my case for denying *Immunity* in the case of authorized threateners. If defensible, my second main claim can be vindicated: an agent's possession of an authority-based justification for causing harm does not, in itself, raise the justificatory burden for defensively harming that agent, compared to if they lacked that justification. Note that this is compatible with the possibility that independent factors *other than their justification* might constrain the permissibility of defense against authorized-threateners. Most obviously, authorized agents seem clearly non-culpable for threatening unjust harm. Though few theorists hold that non-culpable threateners escape liability[39], many accept that a lack of culpability can count against the permissibility of defensive harm to some degree.

Before moving on, it is worth highlighting two implications of the authority-based case for agent-relativity sketched above. Firstly, it suggests that agent-relative reasons can be generated even if it is true that the authority-independent reasons, upon which authoritative commands are based, are entirely agent-neutral. Even in a world populated solely by agent-neutral reasons, authorities may still serve their subjects by issuing commands that enable them to achieve greater conformity with those reasons. But the instrumental reasons created by these commands will be agent-relative. Authoritative

commands that are justified in this way may be understood as a species of 'derivative' agent-relative reasons. These are reasons for action that are specific to certain agents, but whose normative force is derived from their role in enabling that agent to conform to the ultimate, agent-neutral reasons.[40]

Secondly, the Authority View provides a novel argument for both the existence of agent-relative justifications for harm and for the possibility of cases of symmetrically justified harming. Unlike existing agent-relative accounts of permissible harm, the Authority View makes no appeal to considerations of partiality, and so may avoid the standard objections pressed against these accounts.

9. Authority and Constraints

The preceding section sought to show how we can avoid the first horn of our dilemma. An additional argument is required to avoid the second horn: that defensively harming authorized agents is straightforwardly justified in cases such as *Police Officer*, in which resistance seems intuitively impermissible. This conclusion seems forced on us if, as I claim, *Immunity* does not apply in the case of authority-based justifications. This final section aims to provide a plausible and non-*ad hoc* account of why resisting authorized threateners may be impermissible in such cases, which does not appeal to the fact that the threatener is justified (this would simply return us to the first horn of the dilemma).

In the paper so far, I have focused on one important normative consequence of authoritative commands: the creation of a justification for causing harm where none existed antecedently. In order to explain why resisting authorized threateners is sometimes impermissible we need to look at the wider range of normative consequences that authoritative commands can effect. In particular, in addition to providing agents with decisive reasons to perform actions that would otherwise be unjustified, commands may also create decisive reasons to *refrain* from performing actions that would otherwise be *justified*. If the idea of authority-based justifications for harming is defensible, the possibility of authority-based *constraints* should also be.

Once we recognize this additional possibility, we have the resources for explaining why, in cases like *Police Officer*, it may be impermissible to resist an authorized threatener. The key feature of such cases is that the command that harm be caused is addressed to *both* the agent who carries out the harmful action *and* the agent who will suffer the resulting harm. Given this, the command may affect the normative situation of both agents. In particular, the command [Joe be arrested!] may give the police officer a decisive reason to inflict the harm of arrest on Joe *and* give Joe a decisive reason not to exercise his normal right of self-defense. This additional moral power to constrain the use of force may also be justified on service-based grounds (though other forms of justification are possible). For example, an authority's having this power may enable subjects to achieve the coordinative and adjudicative benefits of a system of law.[41] It is this dual exercise of authority, I contend, which explains why resisting authorized-threateners is impermissible in certain cases.[42]

To clarify, the appeal to authority-based constraints is not a tacit reaffirmation of *Immunity*. It's true that both the authorized agent's justification for causing harm and their victim's lack of justification for resisting share the same origin: an authoritative

command issued by a common authority. But these normative consequences are entirely independent of each other. We can see this by noting that agents can be subject to authority-based constraints on resistance even if the aggressors lack any justification. Consider the following:

Invasion: Nation A is facing wholly and clearly unjustified annexation of part of its territory by a more powerful neighbor, Nation B. A's government has service-based authority over its citizens regarding the domain of national defense and, after assessing the expected costs and benefits, commands its citizens not to militarily resist B's agents.

Assume for the sake of argument that resistance by Nation A's citizens is rendered impermissible by the authority's command.[43] This prohibition is clearly not brought about by justification on the part of the aggressors, since they lack any justification whatsoever. Cases such as this demonstrate that authority-based constraints on defense are entirely separable from the moral status of the threatened harm.

10. Conclusion

Reflection on the relationship between the justification of harm and the justification of authority reveals that certain widely held views about the morality of harm and the limits of the obligation to obey require revision. A complete theory of permissible harm will need to make space for both authority-based justifications and authority-based constraints.

---

[1] For written comments on versions of this paper, I am extremely grateful to Yitzhak Benbaji, Christopher Bennett, Cécile Fabre, Adil Ahmed Haque, Jamie Kelly, James Lenman, Jeff McMahan, Jonathan Quong, Massimo Renzo, Daniel Statman, Victor Tadros, and especially Daniel Viehoff. The paper also benefitted from stimulating discussions at a workshop on Victor Tadros' *The Ends of Harm*; The Association for Legal and Social Philosophy Annual Conference; the Oxford War Discussion Group; the inaugural conference of the Stockholm Centre for the Ethics of War and Peace; and seminars at the University of Sheffield and the University of Toronto. I would also like to thank two anonymous reviewers for *Oxford Studies in Political Philosophy* for their extremely perceptive and helpful comments. Work on this paper was supported by the Arts and Humanities Research Council (UK) and the Society for Applied Philosophy.

[2] An exception being a recent pair of articles by Malcolm Thorburn (2008) and John Gardner (2010), though these focus primarily on the relevance of authority to criminal law defences for causing harm. David Estlund (2007) also provides an important discussion, which this paper builds upon.

[3] For a detailed discussion of the nature of these intentions, see Enoch (2011; 2014).

[4] As Helen Frowe puts it, more precisely: "Lesser-evil justifications obtain when one will prevent substantially more harm than one causes, such that the disparity between the harm and the good overrides the deontological presumption against causing harm." (Frowe, 2015, p.274).

[5] For a lucid overview of the agent-neutral/agent-relative distinction, see Ridge (2011).

[6] See, for example, Davis (1984), Quong (2009), and Lazar (2013).

[7] For example, one of the most thorough recent discussions, which "points tantalizingly" towards a "grand unified theory" of permissible harming, consists almost entirely in an analysis of liability and lesser-evil justifications (Rodin, 2011, p.110).

[8] See, for example, Tadros (2011, ch.9), McMahan (2014), and Frowe (2015). For further discussion, see Waldron (2000). Note that *Immunity* does not rule out cases in which two or more agents may be permitted to defensively harm one another as a result of *waiving* their rights against harm (as in boxing matches for example) since these are not cases of rights infringements.

[9] It is worth distinguishing my first claim from an uncontroversial sense in which commands might seem to justify causing otherwise-unjustified harm. To demonstrate, consider a case in which failing to obey a command to cause unjust harm will result in a bad consequence occurring. Perhaps, if disobeyed, the commander will unleash their wrath on innocent people. If this bad consequence is sufficiently grave then the subject may well be justified in acting as commanded. However, in such cases, while the command *results* in reasons for action that justify causing harm, the command *itself* does not create those reasons. Rather, the existence of the command simply affects non-normative facts so as to activate an ordinary lesser-evil justification. For discussion of this distinction, see Estlund (2008, p.118) and Enoch (2014).

[10] For the most influential formulation of the paradox, see Wolff (1970).

[11] On this distinction, see Christiano (2008, pp.261-262).

[12] For explicit defence of this view, see Knowles (2007).

[13] For versions of this view, see McMahan (2009, p.88) and Stilz, (2014, p.333 fn.22).

[14] Some object that this way of justifying authority fails because it cannot account for the idea that having authority necessarily involves a claim right to rule that correlates with the subject's duty to obey (Darwall 2009; 2010). However, it is far from obvious that authority does require such a right, rather than a power to create duties. For (in my view) convincing refutations, see Enoch (2014), Marmor (2011b) and Raz (2010). As these authors point out, in many (perhaps most) cases it seems far more morally attractive to say that the duty to obey an authority is owed to those that the authority's powers are meant to benefit, rather than to the authority itself. To this, let me add that this seems especially plausible when it comes to the specific application of a service account that I argue for, in which the authority's purpose is to enable its subjects to better distribute harms.

[15] An authority may also serve its subject by reducing the burdens of deliberation (Raz, 2009, pp.149-150). I will leave this possibility aside, since this ground of authority is unlikely apply in the contexts I will be discussing.

[16] Though common, this characterisation is an oversimplification, since service accounts can accommodate non-instrumental reasons to obey (see Viehoff, 2011). However, since the cases I discuss are not of this type, I will continue to characterise service accounts as instrumental.

[17] Some argue that the Normal Justification Thesis does not provide a sufficient condition for legitimate authority, because genuine authority must also be conferred by some institutional norm or practice (Marmor, 2011a; 2011b), or at least recognized by some informal social practice (Enoch, 2014). I am unsure whether this is true, but for the purposes of this paper we need not settle the matter, since my specific conclusions about harm do not require that the Normal Justification Thesis be sufficient. Each of these putative necessity conditions are compatible with the view that that service plays a significant role in the justification of authority, whether or not it is sufficient (Enoch, for example, is fairly explicit that his view can be understood as a friendly modification of the service conception). This is all that my argument requires. If it turns out that authority does require institutions or social practices, then my conclusions are accordingly limited to those contexts. However, given that the questions motivating this paper are most likely to arise in precisely those contexts, this doesn't seem particularly worrying.

[18] In addition, pre-emption can also be defended via an argument from double-counting. Since, on a service account, valid commands are wholly grounded the reasons that apply to the subject, subjects cannot simultaneously be subject to a command and the reasons upon which it is based, because these reasons have already been accounted for in producing the command (Raz, 1986, pp.58-59).

[19] In his more recent writings, Raz (2009, p.137) terms this restriction the 'independence condition'. For earlier statements, see Raz (1986, p.57; 1989, p.1180). For detailed discussion, see Tucker (2012).

[20] It is worth pointing out that commands can be invalidated as clear mistakes even if the subject only finds themselves in a position to form the relevant judgment by accident or good fortune.

[21] Thanks to an anonymous referee for pressing me to address this objection.

[22] Some have objected that achieving coordination does not require authorities with the power to impose duties, but only that one course of action be made salient (Green, 1983; 1985). However, as Raz (1989) points out, this seems true only in extremely specific kinds of coordination problems, in which the participants satisfy certain subjective conditions.

[23] In his most recent writings on the topic, Raz seems willing to concede this (Raz, 2010).

[24] This version of the Invalidation Objection thus denies the third step in the four-stage argument for my first claim.

[25] The revised objection is actually broader than the original, since it would also invalidate commands to cause *justified* harm.

[26] I think this response also suggests how the Authority View can respond to a more general class of objections to service accounts. These objections maintain that service accounts fail to give a plausible general account of political authority, because they give insufficient attention to the role that *procedural*

considerations play in justifying authority (such as fairness, democracy, public reason, etc.), focusing instead on the value of the *outcome* of following the authority's directives (see, for example, Christiano (2004); Hershovitz (2003); Quong (2011); Waldron (1999)). While I think these objections are mistaken and that service accounts are flexible enough to accommodate procedural values (see Viehoff, 2011; 2014), the point I want to highlight is that when it comes to the distribution of serious harms, considerations of outcome are intuitively paramount. So, when restricted specifically to harm-apt domains, the case for justifying authority in terms of service is at its most compelling.[26] Some critics seem wiling to concede this. For example, Jonathan Quong, who is otherwise critical of service accounts, agrees that they provide a convincing account of the justification of authority within fairly narrow domains of morally important reasons, such as those constituted by our basic rights and duties *vis-à-vis* one another (Quong, 2011, ch.4). This will presumably include the moral reasons governing the distribution of harm.

[27] For the characterisation of agent-neutrality in terms of common aims, see Parfit (1984, p.27).

[28] For example, Victor Tadros (2011, ch.9) defends a thoroughly agent-neutral view of permissible harming in general and explicitly appeals to this view in order to reject the possibility of symmetrically justified harming (with the exception of cases in which the conflict is itself agent-neutrally valuable, such as in certain sporting contests).

[29] This is not to deny that there are alternative explanations of the intuition that it is impermissible for the innocent arrestee to resist that are compatible with *Completeness*. For example, defence may be futile or counter-productive, given that other police officers will act to make the arrest even if the initial arresting officer is successfully resisted. In my view, such explanations are unsatisfactorily contingent. Thanks to James Lenman and Jeff McMahan for raising this point.

[30] See Estlund (2007) for further discussion of this sort of case.

[31] The right to resist all forms of military aggression has recently come under sustained criticism, so it is not necessarily counter-intuitive to claim that resistance may be unjustified in a case like the one described above (see, especially, Rodin, 2014). However, what is counter-intuitive, even on the most pacifistic views, is the conclusion that resistance may be unjustified *because* the aggressors act with justification. Hence, the oddity of the conclusion generated by applying *Immunity* in the *Combatants* case can be appreciated regardless of one's position on the right to resist military aggression.

[32] On the idea that authority is 'piecemeal' in this way, see Raz (1986, p.71 & p.80.)

[33] This objection has been put to me, independently, by Massimo Renzo and Yitzhak Benbaji.

[34] On the assumption that the mistaken directive is not also binding on me. This is an important caveat, which I discuss in the next section.

[35] Thanks to two anonymous referees for helping me see this.

[36] This slogan is Jeff McMahan's. Note that the doctrine is weaker than *Immunity*, since it is compatible with it being permissible to defensively harm justified threateners on grounds other than liability.

[37] An objection made by Steinhoff (2008) and Rodin (2011), who reject the doctrine. For responses, see McMahan (2008).

[38] Though I cannot argue for it here, I suspect that this may also be true of agent-relative justifications more generally, or at least those grounded in prerogatives to show partiality towards one's own interests (see Quong, 2009). These justifications can also be characterized as granting defenders a permission to act against the just distribution of harm. So if, as McMahan claims, liability tracks the just distribution of harm, the prerogative should not defeat liability.

[39] For a notable exception, see Lazar (2009).

[40] For discussion of this type of reason see, Gardner (2007, p.65) and Hooker (2000, p.110).

[41] For a discussion of how authorities may serve their subjects on grounds of adjudication, see Viehoff (2011).

[42] Of course, authority-based constraints will not always be decisive. For example, in cases where the authority's power to impose this constraint is grounded in its enabling subjects to better comply with their reasons to coordinate with others and adjudicate disagreements impartially, the constraint may be overridden by countervailing considerations that fall outside this domain, such as the costs the subject would have to bear by obeying. My argument can thus accommodate the intuition that if the harm faced by Joe were more serious – long term imprisonment for example – then he may be morally justified in resisting, despite being commanded not to.

[43] For a different argument for the possibility of such cases, see Stilz (2014).

Works Cited

Christiano, T. (2004) 'The Authority of Democracy', *Journal of Political Philosophy*, 12(3), pp.266–90.

Christiano, T. (2008) *The Constitution of Equality*. Oxford: Oxford University Press.

Darwall, S. (2009) 'Authority and Second-Personal Reason for Acting', in Sobel, D. and Wall, S. (eds) *Reasons for Action*. Cambridge: Cambridge University Press, pp.134-154.

Darwall, S. (2010) 'Authority and Reasons: Exclusionary and Second Personal', *Ethics* 120(2), pp.257-278.

Davis, N.A. (1984) 'Abortion and Self-Defense', *Philosophy and Public Affairs*, 13(2), pp.175-207.

Enoch, D. (2011) 'Giving Practical Reasons', *Philosopher's Imprint*, 11(4), pp.1-21.

Enoch, D. (2014) 'Authority and Reason-Giving', *Philosophy and Phenomenological Research*, 89(2), pp.296-332.

Estlund, D. (2007) 'On Following Orders in an Unjust War', *Journal of Political Philosophy*, 15(2), pp.213-234.

Estlund, D. (2008) *Democratic Authority*. Princeton: Princeton University Press.

Frowe, H. (2015) 'Claim Rights, Duties and Lesser-Evil Justifications', *Proceedings of the Aristotelian Society* 89(1), pp.267-285.

Gardner, J. (2007) *Offences and Defences*. Oxford: Oxford University Press.

Gardner, J. (2010) 'Justification Under Authority', *Canadian Journal of Law and Jurisprudence*, 23(1), pp.73-98.

Green, L. (1983) 'Law, Coordination, and the Common Good', *Oxford Journal of Legal Studies*, 3(3), pp.299-324.

Green, L. (1985) 'Authority and Convention', *The Philosophical Quarterly*, 35(141), pp.329-346.

Hershovitz, S. (2003) 'Legitimacy, Democracy, and Razian Authority', *Legal Theory*, 9(3), pp.201–220.

Hooker, B. (2000) *Ideal Code, Real World*. Oxford: Oxford University Press.

Hosein, A. (2014) 'Are Justified Aggressors a Threat to the Rights Theory of Self-Defense?', in Frowe, H. and Lang, G. (eds) *How We Fight*. Oxford: Oxford University Press, pp.87-103.

Knowles, D. (2007) 'The Domain of Authority', *Philosophy*, 82(1), pp.23-43.

Lazar, S. (2009) 'Responsibility, Risk, and Killing in Self-Defense', *Ethics*, 119(4), pp.699-728.

Lazar, S. (2013) 'Associative Duties and the Ethics of Killing in War', *Journal of Practical Ethics*, 1(1), pp.6-51.

Mapel, D. (2010) 'Moral Liability to Defensive Harm and Symmetrical Self-Defense', *Journal of Political Philosophy*, 18(2), pp.198-217.

Marmor, A. (2011a) 'The Dilemma of Authority', *Jurisprudence*, 2(1), pp.121-141.

Marmor, A. (2011b) 'An Institutional Conception of Authority', *Philosophy and Public Affairs*, 39(3), pp.238-261.

McMahan, J. (2008) 'Justification and Liability', *Journal of Political Philosophy*, 16(2), pp.227-244.

McMahan, J. (2009) *Killing in War*. Oxford: Oxford University Press.

McMahan, J. (2014) 'Self-Defense Against Justified Threateners', in Frowe, H. and Lang, G. (eds) *How We Fight*. Oxford: Oxford University Press, pp.104-137.

Parfit, D. (1984) *Reasons and Persons*. Oxford: Oxford University Press.

Quong, J. (2009) 'Killing in Self-Defense', *Ethics*, 119(3), pp.507-537.

Quong, J. (2011) *Liberalism Without Perfection*. Oxford: Oxford University Press.

Raz, J. (1986) *The Morality of Freedom*. Oxford: Clarendon Press.

Raz, J. (1989) 'Facing Up: A Reply', *Southern California Law Review*, 62, pp.1153-1235.

Raz, J. (2009) *Between Authority and Interpretation*. Oxford: Oxford University Press.

Raz, J. (2010) 'On Respect, Authority and Neutrality: A Response', *Ethics*, 120(2), pp.279-301.

Ridge, M. (2011) 'Reasons for Action: Agent-Neutral vs. Agent-Relative', in Zalta, E.N. (ed) *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/win2011/entries/reasons-agent/>

Rodin, D. (2011) 'Justifying Harm', *Ethics*, 122(1), pp.74-110.

Rodin, D. (2014) 'The Myth of National Defence', in Fabre, C. and Lazar, S. (eds) *The Morality of Defensive War*. Oxford: Oxford University Press, pp.69-89.

Smith, M.N. (2013) 'Political Obligation and the Self', *Philosophy and Phenomenological Research*, 86(2), pp.347-375.

Steinhoff, U. (2008) 'Jeff McMahan on the Moral Equality of Combatants', *Journal of Political Philosophy*, 16(2), 220-226.

Stilz, A. (2014) 'Authority, Self-Determination and Community in Cosmopolitan War', *Law and Philosophy*, 33(3), pp.309-335.

Tadros, V. (2011) *The Ends of* Harm. Oxford: Oxford University Press.

Thorburn, M. (2008) 'Justifications, Powers, and Authority', *Yale Law Journal* 117, pp.1070-1130.

Tucker, A. (2012) 'The Limits of Razian Authority', *Res Publica*, 18(3), pp.225-240.

Viehoff, D. (2011) 'Procedure and Outcome in the Justification of Authority', *Journal of Political Philosophy*, 19(2), pp.248-259.

Viehoff, D. (2014) 'Democratic Equality and Political Authority', *Philosophy and Public Affairs*, 42(4), pp.337-375.

Waldron, J. (2000) 'Self-Defense: Agent-Neutral and Agent-Relative Accounts', *California Law Review*, 88, pp.711-750.

Waldron, J. (1999) *Law and Disagreement*. Oxford: Oxford University Press.

Wolff, R.P. (1970) *In Defense of Anarchism*. New York: Harper.